

生物系の数理科学 第10回

有限集団における遺伝的浮動

山口諒

1 確率性とヘテロ接合度の減少

集団遺伝学において、遺伝的浮動 (genetic drift) は、ヘテロ接合度 (heterozygosity) の減少に重要な影響を与える。ヘテロ接合度 H は、ランダムに選んだ2つの対立遺伝子が異なる確率を表す。これに対して、ホモ接合度 (homozygosity) G は、それらが同一である確率であり、

$$H = 1 - G$$

である。

1.1 ヘテロ接合度の減少率

遺伝的浮動の影響を受けた1世代後のホモ接合度 G' は、以下の式で表される：

$$G' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G.$$

この式の導出について説明する。 G' は、次世代で2つのランダムに選んだ対立遺伝子が同一である確率を表す。この確率は、2つの異なるプロセスによって決定される。

- **右辺第1項** $\frac{1}{2N}$ は、次世代でランダムに選ばれた2つの対立遺伝子が、同一の祖先遺伝子を持つ確率である。各個体には2つの対立遺伝子があり、個体数が N の場合、1つの対立遺伝子が親世代からランダムに選ばれる確率は $1/2N$ となる。
- **右辺第2項** $\left(1 - \frac{1}{2N}\right) G$ は、ランダムに選んだ2つの対立遺伝子が異なる祖先を持っていたものの、それらの祖先遺伝子が同一であった確率を示す。前世代で2つの異なる遺伝子が選ばれた確率は $1 - 1/2N$ であり、さらにそれらが同一遺伝子であった確率は G となる。

この2つの確率の和が G' となる。この式を H で表すために、 $H' = 1 - G'$ を計算すると、

$$H' = \left(1 - \frac{1}{2N}\right) H.$$

この関係を用いて、1世代あたりの変化量を $\Delta_N H$ と定義すると、

$$\Delta_N H = H' - H = -\frac{1}{2N} H.$$

すなわち、ヘテロ接合度は世代ごとに $1/2N$ の割合で減少する。

1.2 ヘテロ接合度の時間変化

ヘテロ接合度の時間経過に関する一般式は、次の漸化式を帰納的に解くことで得られる：

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t.$$

これは指数関数的減衰を示し、十分に大きな N に対しては、

$$H_t \approx H_0 e^{-t/2N}.$$

となる。

1.3 ヘテロ接合度が半減する時間

ヘテロ接合度が初期値 H_0 の半分に減少するまでの時間 $t_{1/2}$ は、以下の方程式を解くことで求められる：

$$\frac{H_0}{2} = H_0 \left(1 - \frac{1}{2N}\right)^t.$$

これを対数変換すると、

$$t_{1/2} = \frac{-\ln(2)}{\ln(1 - 1/2N)}.$$

近似式 $\ln(1 + x) \approx x$ を用いると、

$$t_{1/2} \approx 2N \ln 2.$$

これは、集団サイズが大きいほど遺伝的浮動の影響が小さく、ヘテロ接合度の減少が遅くなることを意味する。

1.4 突然変異を考慮したモデル

突然変異率を u とすると、1世代後の G' は以下の式で与えられる：

$$G' = (1 - u)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G \right].$$

この式を展開し、 $(1 - u)^2 \approx 1 - 2u$ を用いて近似すると、

$$G' \approx (1 - 2u) \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G \right].$$

さらに、 $H = 1 - G$ を代入し整理すると、

$$H' \approx \left(1 - \frac{1}{2N}\right) H + 2u(1 - H).$$

1 世代ごとの変化量 ΔH は、

$$\Delta H \approx -\frac{1}{2N}H + 2u(1 - H).$$

平衡状態では $\Delta H = 0$ となるため、ヘテロ接合度の平衡値 H^* は、

$$H^* = \frac{4Nu}{1 + 4Nu}.$$

となる。

2 ライト・フィッシャー過程と二項分布

2.1 ライト・フィッシャー過程の概要

ライト・フィッシャー (Wright-Fisher) 過程は、**有限集団**における遺伝的浮動 (genetic drift) を定量的に扱うための代表的なモデルである。集団サイズを二倍体で N 個体とすると、合計で $2N$ 本のアリル (遺伝子コピー) が存在する。このモデルでは、次のような仮定を置く。

- **離散世代**: 世代から世代へと個体が一斉に切り替わる。
- **ランダムサンプリング**: 次世代を構成するアリルは、現世代のアリルを等確率で「抜き取り・置き戻し」方式により選ぶ (**独立反復試行**)。
- **集団サイズ一定**: 各世代の個体数は常に N と固定される。

ある対立遺伝子 A が、現世代でアリル頻度 p を持っているとは仮定する。すると、次世代でこの遺伝子 A が何本出現するかは、**各サンプリングが独立で成功確率 p の試行が $2N$ 回行われるイメージ**でモデル化できる。ここで登場するのが**二項分布**である。

2.1.1 次世代のアリル頻度と二項分布

次世代におけるアリル A のコピー数 X は、成功確率 p 、試行回数 $2N$ の二項分布に従うと想定される。すなわち

$$X \sim \text{Binomial}(2N, p).$$

したがって、「次世代で A がちょうど k 本存在する」確率は

$$P(X = k) = \binom{2N}{k} p^k (1 - p)^{2N-k}, \quad k = 0, 1, \dots, 2N.$$

ここで、アリル頻度としては

$$\hat{p} = \frac{X}{2N}$$

が次世代の A の頻度となる。有限集団では X がランダムに変動するため、 \hat{p} も確率的に揺らぐ。これが**遺伝的浮動**であり、集団サイズ N が小さいほどその変動幅は大きくなる。

2.2 二項分布の基礎

上述のようにライト・フィッシャー過程の鍵となるのが二項分布である。二項分布は、「 n 回の独立試行」で「成功確率 p の事象」が「何回成功するか」を表す分布だと考えられる。

2.2.1 定義と導出

二項分布 (Binomial distribution) とは、試行回数 n および成功確率 p をパラメータとする確率分布で、成功回数を表す確率変数 X が以下を満たす：

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

ここで、 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ は二項係数である。

- p^k : k 回の成功確率をすべて掛け合わせたもの。
- $(1-p)^{n-k}$: 残り $n-k$ 回の失敗確率の積。
- $\binom{n}{k}$: どの試行が成功で、どの試行が失敗かをアレンジする組合せの数。

2.2.2 平均と分散

二項分布 $X \sim \text{Binomial}(n, p)$ とは、 n 回の独立なベルヌーイ試行 (成功確率 p) における成功回数を表す確率分布である。ここでは、二項分布の平均 (期待値) と分散の導出を示す。

導出 1：指示変数を用いた導出 二項分布 X は、独立なベルヌーイ乱数

$$I_1, I_2, \dots, I_n \quad (\text{それぞれ Bernoulli}(p))$$

の和として

$$X = I_1 + I_2 + \dots + I_n$$

と書ける。ここで I_k は、「 k 回目の試行で成功すれば 1, 失敗すれば 0 をとる」指示変数である。

1. $E[X] = np$

まず期待値を求める。期待値の線形性より、

$$E[X] = E\left[\sum_{k=1}^n I_k\right] = \sum_{k=1}^n E[I_k].$$

それぞれの I_k は成功確率 p のベルヌーイ乱数なので、 $E[I_k] = p$ である。よって、

$$\sum_{k=1}^n E[I_k] = \sum_{k=1}^n p = np.$$

したがって、

$$E[X] = np.$$

直観的にも、 n 回中およそ p の割合で成功することを期待するため、平均は np になると考えられる。

2. $V[X] = n p (1 - p)$

次に分散を求める。 X は独立なベルヌーイ乱数の和なので、分散も加法的に求められる。すなわち、

$$V[X] = V\left(\sum_{k=1}^n I_k\right) = \sum_{k=1}^n V[I_k].$$

ベルヌーイ乱数 I_k の分散は $p(1 - p)$ であるから、

$$\sum_{k=1}^n V[I_k] = \sum_{k=1}^n p(1 - p) = n p (1 - p).$$

よって、

$$V[X] = n p (1 - p).$$

これは、成功確率 p と失敗確率 $1 - p$ の「バランス」によって決まり、 p が 0.5 に近いほど分散が大きくなる（成功と失敗のばらつきが大きい）ことを示す。

(補足説明) ベルヌーイ乱数の分散 各 I_k はベルヌーイ乱数（または指示変数）であり、次のように 0 と 1 の値しかとらない：

$$I_k = \begin{cases} 1 & (\text{成功, 確率 } p), \\ 0 & (\text{失敗, 確率 } 1 - p). \end{cases}$$

このとき、ベルヌーイ乱数の分散が $p(1 - p)$ となる理由を以下に示す。

一般に、確率変数 Y の分散は

$$V[Y] = E[Y^2] - (E[Y])^2$$

で与えられる。ベルヌーイ乱数 I_k は値が 0 または 1 のみなので

$$I_k^2 = I_k$$

が常に成り立つ。そこで期待値を計算すると

$$E[I_k] = 1 \cdot p + 0 \cdot (1 - p) = p, \quad E[I_k^2] = E[I_k] = p.$$

よって

$$V[I_k] = E[I_k^2] - (E[I_k])^2 = p - p^2 = p(1 - p).$$

したがって、それぞれの I_k の分散は $p(1 - p)$ となる。

2.3 遺伝的浮動と分散の解釈

ライト・フィッシャー過程での次世代における A アリルコピー数 X は、 $\text{Binomial}(2N, p)$ に従う。すると、

$$E[X] = 2N p, \quad V[X] = 2N p (1 - p).$$

次世代のアリル頻度は $\hat{p} = \frac{X}{2N}$ となるので、

$$E[\hat{p}] = \frac{E[X]}{2N} = p,$$

$$V(\hat{p}) = \frac{V[X]}{(2N)^2} = \frac{2N p(1-p)}{(2N)^2} = \frac{p(1-p)}{2N}.$$

したがって、平均的には同じ頻度 p を維持するものの、次世代では**分散**が $\frac{p(1-p)}{2N}$ だけ生じる。これが**遺伝的浮動**と呼ばれる確率的揺らぎの源泉であり、集団サイズ N が小さいほどその効果が顕著になる。最終的に固定（ある遺伝子が 100% に達する）や消失（0% になる）が、適応度の差が無い状況でもランダムに起こりうる。